Lois de probabilité variables discrètes

Type de loi	Généralités	Propriétés	Fonction de probabilité	Conditions	Programme SAS
Loi binomiale	Épreuve de Bernoulli: *succès* ou *échec* sur un *essai* Processus de Bernoulli: *succès* ou *échec* sur n *essais* Loi binomiale: nombre de *succès* sur n *essais*? Contexte épidémiologique: • essai : observation (ex.: participant dans une étude) • succès : évènement (ex.: décès) • proportion : # d'évènements / # d'observations	Paramètres: π, n E(X) = nπ V(X) = nπ(1- π) Domaine: 0 ≤ X ≤n	$Pr(X=x) = C_n^x \pi^x (1-\pi)^{n-x}$ X : nombre d'évènements n : nombre d'observations π : probabilité d'un évènement C : combinaison: $C_n^x = \frac{n!}{x!(n-x)!}$	-Chaque observation ne peut entraîner qu'un évènement ou un non-évènement; -Les observations sont indépendantes, c'est-à-dire que la probabilité π_i d'un évènement à la $i^{ième}$ observation ne dépend pas des résultats des autres observations; -Les probabilités π_i sont les mêmes à chaque observation, c'est-à-dire que $\pi_i = \pi$ pour tous les i.	PROBBNML (π,n, x)
Loi de Poisson	Expérience de Poisson: nombre de *succès* pour une unité de temps (dt) Processus de Poisson: nombre de *succès* pour plusieurs unités de densité (dt) Loi de Poisson: combien de *succès* pour plusieurs dt? Contexte épidémiologique: • succès : événement • dt : une unité de personne-temps • taux : nombre d'évènements pour une unité de personne-temps • Les personnes-temps sont une fonction du nombre de participants observés et du temps	Paramètres: λ E(X) = λ V(X) = λ Domaine: 0 ≤ X ≤ ?	$P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$ X : nombre d'évènements λ : nombre d'évènements attendus pour la période d'observation selon τ : $\lambda = \tau \times ds$	-La réalisation des événements dans un intervalle de temps est indépendante de ce qui s'est passé dans les intervalles précédents; -La probabilité que se produise exactement un événement dans un intervalle est proportionnelle à la longueur de cet intervalle; -La probabilité que se produise plus d'un événement dans l'intervalle élémentaire est négligeable.	POISSON (λ, x)
Loi hypergéométrique	Loi hypergéométrique: nombre de *succès* pour un échantillon de n *essais* issue d'une population de N *essais* Fraction d'échantillonnage (n/N) grande Contexte épidémiologique: • succès : évènement • essai : observation • cote : nombre d'évènements/nombre de non-évènements	Paramètres: N, M_1, n $E(X) = \mu = \sum x_i P(x_i) = \frac{n M_1}{N}$ V(X) = $\frac{n M_1 M_0 (N-n)}{N^2 (N-1)}$ Domaine: Max $(0,n-M_0) \leq X \leq$ Min(n,M_1)	$P(X=x) = \frac{C_{M_1}^x \times C_{M_0}^{n-x}}{C_N^n}$ $= \frac{M_1! M_0! n! (N-n)!}{x!(M_1-x)!(n-x)![M_0(n-x)]!N!}$ x : nombre d'évènements dans l'échantillon n : nombre d'observations dans l'échantillon M_1 : nombre d'évènements dans la population M_0 : nombre de non-évènements dans la population N : nombre d'observations dans la population	-Chaque observation ne peut entraîner qu'un évènement ou un non-évènement; -Sélections ne sont pas remplacées; en d'autres mots les observations sont dépendantes, c'est-à-dire que la probabilité d'un évènement à la $i^{ième}$ observation dépend des résultats des autres observations; -L'échantillon doit être aléatoire	PROBHYPR (N, M_1, n, x)

- P(X = x) = P(X ≤ x) – P(X ≤ (x - 1))

Carin Ahouada. Médecin Interniste béninois. Etudiant Maitrise en Epidémiologie. Université Laval. 2016-2018

Lois de probabilité variables continues

Type de loi	Utilisation des lois	Propriétés	Fonction de probabilité	Conditions	Programme SAS
Loi normale	Mesures de position (moyenne) et dispersion (variance), tests statistiques, et intervalles de confiance pour des variables continues gaussiennes Tests statistiques et intervalles de confiance sur variables non-gaussiennes si n est assez élevé (ex.: valeur-p ou IC *asymptotiques* pour des taux ou proportions)	Distribution de densité de probabilité $P(X = x) \approx 0$ $\Sigma_x P(X = x) = 1$	Loi normale centrée réduite À l'aide d'une transformation, toutes les variables continues gaussiennes peuvent être décrites à l'aide d'une même loi de probabilité Les probabilités peuvent ensuite être obtenues à l'aide d'un tableau statistique Paramètres: $X \sim N(\mu_x, \sigma_x^2)$ $Z = \frac{X - \mu_X}{\sigma_X}$ $Z \sim N(0, 1)$	X suit une distribution normale Indépendance des observations	Loi normale: Pour calculer P (Z ≤ z): PROBNORM(z) Pour obtenir z à partir de P(Z ≤ z): PROBIT(P)
Loi de Student	Tests statistiques et intervalles de confiance pour des variables continues (ex.: comparaison de moyennes) lorsque X est gaussienne et que n < 30	La loi de Student tend vers la loi normale lorsque n ≥ 30	Paramètres : t, k $t = \frac{(\overline{X} - \mu)}{s/\sqrt{n}}$ $k = n - 1$	Loi d'échantillonnage (à utiliser si n est petit)	Loi de Student (k degrés de liberté): Pour calculer P(T ≤ t): PROBT(t, k) Pour obtenir t à partir de P(T ≤ t): TINV(P)
Loi du Khi-carré	Tests statistiques *asymptotiques* pour comparer deux ou plusieurs taux ou proportions		Paramètres: χ^2, k Extension de la loi normale Si $Z \mapsto N(0,1)$, alors $Z^2 \mapsto \chi_1^2$		Loi du khi carré* (k degrés de liberté): Pour calculer P(X ≤ x): PROBCHI(x, k)

SAS calcule la probabilité **P(X ≤ x)**

Outils graphiques: • Histogramme • Boîte à moustaches • Tracé normalisé	Test statistique: Test de Shapiro-Wilk • H_0: distribution observée = distribution normale • H_1: distribution observée ≠ distribution normale	Dans SAS: PROC UNIVARIATE normal plots; VAR *var;* HISTOGRAM *var;* RUN;

Carin Ahouada. Médecin Interniste béninois. Etudiant Maitrise en Epidémiologie. Université Laval. 2016-2018

Analyse de moyenne Tests paramétriques

	1 échantillon	2 échantillons indépendants	2 échantillons dépendants	Plus de 2 échantillons
Mesures d'association	$m = \Sigma y_i / n$ Intervalle de confiance: $m \pm z_{\alpha/2} s_y/\sqrt{n}$ m : moyenne observée $z_{\alpha/2}$: percentile de la loi normale selon α (si n<30, on doit utiliser $t_{\alpha/2,n-1}$) s_y : écart-type de la variable Y n : nombre d'observations dans l'échantillon N.B.: $\sigma_y/\sqrt{n} = \sigma_m$	Différence de moyennes (ou moyenne de différences): $dm = m_2 - m_1$ Intervalle de confiance : $dm \pm z_{\alpha/2} s_{dm}$ $dm \pm t_{(n1+n2)-2,\alpha/22} s_{dm}$ lorsque $(n_1+n_2) < 30$ Le calcul de s_{dm} dépend de l'égalité ou non des variances **Si les variances sont égales:** $s_{dm}^2 = s_Y^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)$ où S_Y^2 est calculé à partir de S_{Y1}^2 et S_{Y2}^2: $s_Y^2 = \frac{(n_1-1)s_{Y1}^2 + (n_2-1)s_{Y2}^2}{(n_1-1)+(n_2-1)}$ **Si les variances sont différentes:** $s_{dm}^2 = \left(\frac{s_{Y1}^2}{n_1} + \frac{s_{Y2}^2}{n_2}\right)$ degrés de liberté (correction de Satterthwaite) est obtenu par : $v = \frac{(w_1+w_2)^2}{w_1^2/(n_1-1) + w_2^2/(n_2-1)}$ où $w_i = \frac{s_{Yi}^2}{n_i}$	Mesure d'association: $m_D = \sum d_i / n = m_2 - m_1 = dm$ Intervalle de confiance $m_D \pm z_{\alpha/2} \frac{s_D}{\sqrt{n}}$ Note: $z_{\alpha/2}$ est le percentile de la loi normale associé à α/2 Utiliser $t_{n-1,\alpha/2}$ si n (nombre de paires)<30 $s_{mD} = s_D/\sqrt{n}$	
Tests Test bilatéral	$H_0 : m = \mu$ $H_1 : m \neq \mu$ α = 0.05 $z/t = \frac{(m-\mu)}{s/\sqrt{n}}$ Valeur-p = P (Z ≥ |z|) + P(Z ≤ -|z|) selon la loi normale ou P(T ≥ |z|) + P(T ≤ -|z|) selon la loi de Student Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p < α, on rejette H_0	$H_0 : dm = 0$ $H_1 : dm \neq 0$ α = ? $z/t = \frac{dm}{s_{dm}}$ (attention au calcul de s_{dm}) Selon la loi normale, valeur-p = P(Z ≥ z) + P(Z ≤ -z) ou selon la loi de Student, valeur-p = P(T ≥ t) + P(T ≤ -t) Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p < α, on rejette H_0	$H_0 : m_D = 0$ $H_1 : m_D \neq 0$ α = ? $z/t = \frac{m_D - 0}{\left(\frac{s_D}{\sqrt{n}}\right)} = \frac{m_D}{\left(\frac{s_D}{\sqrt{n}}\right)}$ Selon la loi normale, valeur-p = P (Z ≥ z) + P(Z ≤ -z) ou selon la loi de Student, valeur-p = P(T ≥ t) + P(T ≤ -t) Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p < α, on rejette H_0	ANOVA Test de Fisher global **Hypothèse statistique:** $H_0 : m_1 = m_2 = \ldots = m_j$ H_1 : Au moins deux moyennes diffèrent Interprétation: Si on accepte H_0 (valeur-p ≥ α), on conclut que toutes les moyennes sont statistiquement équivalentes Si on rejette H_0 (valeur-p < α), on conclut qu'au moins deux des moyennes sont statistiquement différentes **ANOVA** **Corrections pour comparaisons multiples** **Tukey:** convient quand les tailles de chacun des groupes (n) sont égales **Scheffe**: convient si n varie d'un groupe à l'autre (cette correction est plus conservatrice que Tukey) **Bonferonni**: trop conservateur si le nombre de comparaisons est élevé On peut aussi réduire le nombre de comparaisons en spécifiant seulement celles qui nous intéressent

Carin Ahouada. Médecin Interniste béninois. Etudiant Maitrise en Epidémiologie. Université Laval. 2016-2018

Analyse de moyenne Tests paramétriques				
	1 échantillon	**2 échantillons indépendants**	**2 échantillons dépendants**	**Plus de 2 échantillons**
Taille de l'échantillon	**Selon une précision *i*** $n = \left[z_{\alpha/2} \frac{\sigma_Y}{i} \right]^2$ n : nombre d'observations requis $z_{\alpha/2}$: percentile de la loi normale associé à α/2 σ_y : écart-type de la variable Y i : précision désirée (moitié de la largeur d'un intervalle de confiance) **selon une une différence Δ** $n = \frac{\sigma_Y^2 \ (z_{\alpha/2} + z_\beta)^2}{\Delta^2}$ n : nombre d'observations requis $z_{\alpha/2}$: percentile de la loi normale associé à α/2 (Z_α pour un test unilatérale) z_β : percentile de la loi normale associé à β (1-β est la puissance désirée) σ_y^2 : variance de la variable Y Δ : différence entre m et μ qu'on souhaite détecter	**Selon une précision *i*** $n_1 = \frac{z_{\alpha/2}^2 \sigma_Y^2}{i^2} \left(\frac{k+1}{k} \right)$ n_1 : nombre d'observations requis dans le groupe 1 $z_{\alpha/2}$: percentile de la loi normale associé à α/2 σ_Y : écart-type de la variable Y i : précision désirée pour dm (moitié de la largeur de l'intervalle de confiance) k : n_2 / n_1 **selon une différence Δ** $n_1 \geq \frac{\sigma^2 (z_\alpha + z_\beta)^2}{\Delta^2} \left[\frac{k+1}{k} \right]$ n_1 : nombre d'observations requis dans le groupe 1 $z_{\alpha/2}$: percentile de la loi normale associé à α/2 (Z_α pour un test unilatéral) z_β : percentile de la loi normale associé à β (1-β est la puissance désirée) σ_Y^2 : variance de la variable Y Δ : différence de moyennes qu'on souhaite détecter k : n_2 / n_1	**Pour détecter une différence moyenne d'au moins Δ** $n_1 \geq \frac{\sigma^2 (z_\alpha + z_\beta)^2}{\Delta^2} \left[\frac{k+1}{k} \right]$ n' : nombre de paires d'observations requises $Z_{\alpha/2}$: percentile de la loi normale associé à α/2 (Z_α pour un test unilatérale) Z_β : percentile de la loi normale associé à β (1-β est la puissance désirée) σ_D^2 : variance de la variable D Δ : différence moyenne qu'on souhaite détecter N.B. Si on ne connait pas σ_D^2 , on peux le calculer à partir de σ_Y^2 et du coefficient de corrélation ρ_{Y1Y2} : $\sigma_D^2 = 2\sigma_Y^2 (1 - \rho_{Y1,Y2})$	
Puissance		**une différence de Δ entre deux moyennes** $z_\beta = \frac{\lvert\Delta\rvert}{\sigma_Y} \sqrt{\frac{n_1 \times n_2}{n}} - z_{\alpha/2}$ z_β : percentile de la loi normale associé à β (1-β est la puissance obtenue) n_1 : nombre d'observations dans le groupe 1 n_2 : nombre d'observations dans le groupe 2 n : nombre total d'observations $z_{\alpha/2}$: percentile de la loi normale associé à α/2 (Z_α pour un test unilatéral) σ_Y : écart-type de la variable Y Δ : différence de moyennes qu'on souhaite détecter	**Pour détecter une différence moyenne d'au moins Δ** $z_\beta = \frac{\lvert\Delta\rvert \sqrt{n'}}{\sigma_D} - z_{\alpha/2}$ z_β : percentile de la loi normale associé à β (1-β est la puissance obtenue) n' : nombre de paires $z_{\alpha/2}$: percentile de la loi normale associé à α/2 (z_α pour un test unilatéral) σ_D : écart-type de la variable D (on peut le calculer à partir de σ_Y^2 et ρ_{Y1Y2}) Δ : différence moyenne qu'on souhaite détecter	

Carin Ahouada. Médecin Interniste béninois. Etudiant Maitrise en Epidémiologie. Université Laval. 2016-2018

Analyse de moyenne Tests paramétriques				
	1 échantillon	**2 échantillons indépendants**	**2 échantillons dépendants**	**Plus de 2 échantillons**
SAS	**Moyenne et IC** PROC UNIVARIATE cibasic; VAR *y*; RUN; **comparaison avec moyenne théorique** PROC UNIVARIATE mu0=μ;VAR *y*; RUN; **Taille selon la précision désirée** PROC POWER; ONESAMPLEMEANS ci=t halfwidth = *i* stddev = σ ntotal = . probwidth = alpha = α; RUN; **Taille selon une différence clinique** PROC POWER; ONESAMPLEMEANS nullmean = μ mean = *m* stddev = σ ntotal = . alpha = α sides = *1 or 2* power = *1*-β; RUN;	**PROC TTEST**; CLASS *X*; VAR *Y*; **RUN**; où X désigne les groupes; et Y, la variable d'intérêt **taille selon une difference** PROC POWER; TWOSAMPLEMEANS meandiff = Δ stddev = σ_Y ntotal =. power = *1*-β alpha = α sides = *1 or 2*; RUN; **Puissance** PROC POWER; TWOSAMPLEMEANS meandiff = Δ stddev = σ_Y ntotal =n power = . alpha = α sides = *1 or 2*; RUN;	**PROC TTEST**; VAR *D*; **RUN**; ou pour avoir les résultats de tests statistiques paramétrique et non-paramétrique ainsi que des tests de normalité dans la même sortie: **PROC UNIVARIATE** NORMAL PLOTS; VAR *D*; **RUN**; **Taille pour détecter une différence moyenne d'au moins Δ** **proc power;** pairedmeans meandiff = m_D pairedstddevs = σ_{Y1} | σ_{Y2} corr = ρ npairs = . alpha = α sides = *1 or 2* power = *1* - β; **RUN**; **Puissance détecter une différence moyenne d'au moins Δ** **proc power;** pairedmeans meandiff = m_D pairedstddevs = σ_{Y1} | σ_{Y2} corr = ρ npairs = n alpha = α sides = *1 or 2* power = .; **RUN**;	**PROC ANOVA**; CLASS X; MODEL *Y* = *X*; MEANS X / tukey scheffe clm cldiff hovtest=levene; **RUN**;
Conditions	**Test paramétrique** -Si n < 30, Y doit suivre une loi normale. Vérifier la normalité de Y si n < 30! -Les observations sont indépendantes. -L'échantillon (population d'étude) est représentatif de la population cible. -Absence d'erreur de mesure. -Absence de biais de confusion	Les observations sont indépendantes. Si dans un (ou les deux) échantillon(s) n < 30, Y doit suivre une distribution normale dans cet (ces) échantillon(s). La population d'étude est représentative de la population cible. Absence d'erreur de mesure. Absence de biais de confusion.	Si le nombre de paires n' < 30, la différence D suit une distribution normale Indépendance des paires d'observations La population d'étude est représentative de la population cible Absence d'erreur de mesure Absence de biais de confusion	Si n < 30 dans un échantillon, Y est gaussienne Variances égales

Carin Ahouada. Médecin Interniste béninois. Etudiant Maitrise en Epidémiologie. Université Laval. 2016-2018

Analyse de moyens tests non paramétriques				
	1 échantillon	2 échantillons indépendants	2 échantillons dépendants	Plus de 2 échantillons
Test statistique bilatéral		**Test de Wilcoxon de somme des rangs** H_0: échantillon$_1$ = échantillon$_2$ H_1: échantillon$_1$ ≠ échantillon$_2$ α = ? $z = t = \frac{W_1 - W_a}{S_{W_1}}$ w_1 : somme des rangs pour un des échantillons w_a : somme attendue, soit $W_a = \frac{n_1(n+1)}{2}$ $S_{W_1} = \sqrt{\frac{n_1 n_2 (n+1)}{12}}$ Selon la loi normale, valeur-p = P(Z ≥ z) + P(Z ≤ -z) ou selon la loi de Student, valeur-p = P(T ≥ t) + P(T ≤ -t) Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p < α, on rejette H_0	**Test des *rangs signés* de Wilcoxon** H_0: échantillon$_1$ = échantillon$_2$ H_1: échantillon$_1$ ≠ échantillon$_2$ α = ? $z = t = \frac{W_p - W_a}{S_{W_p}}$ w_p : somme des rangs des différences positives w_a : somme attendue, soit $W_a = \frac{n'(n'+1)}{4}$ $S_{W_p} = \sqrt{\frac{n'(n'+1)(2n'+1)}{24}}$ Selon la loi normale, valeur-p = P(Z ≥ z) + P(Z ≤ -z) ou selon la loi de Student, valeur-p = P(T ≥ t) + P(T ≤ -t) Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p < α, on rejette H_0	**Test de Kruskall-Wallis** Hypothèse statistique H_0: échantillon$_1$ = échantillon$_2$ = … = échantillon$_j$ H_1: Au moins deux échantillons diffèrent
Conditions		n > 10 dans chacun des échantillons Pas trop de multiplicité, i.e. pas ou peu de rangs ex-aequo Échantillons (population d'étude) représentatifs de la population cible Absence d'erreur de mesure Absence de biais de confusion	n' > 20 Pas trop de multiplicité, i.e. pas ou peu de rangs ex-aequo Échantillons (population d'étude) représentatifs de la population cible Absence d'erreur de mesure Absence de biais de confusion	n > 10 dans chacun des échantillons Pas trop de multiplicité, i.e. pas ou peu de rangs ex-aequo Échantillons (population d'étude) représentatifs de la population cible Absence d'erreur de mesure Absence de biais de confusion
Sas		**PROC NPAR1WAY** wilcoxon; CLASS X; VAR Y; **RUN**;	**PROC UNIVARIATE** normal plots; VAR D; **RUN**; Permet d'obtenir les tests statistiques paramétrique et non paramétrique, ainsi que des diagnostiques de normalité	**PROC NPAR1WAY** wilcoxon; CLASS X; VAR Y; **RUN**;

Carin Ahouada. Médecin Interniste béninois. Etudiant Maitrise en Epidémiologie. Université Laval. 2016-2018

Analyse de proportions

Mesures d'association

1 proportion

Estimation d'une proportion

Intervalle de confiance asymptotique:

$$p \pm z_{\alpha/2}\sqrt{\frac{pq}{n}}$$

p – proportion de malades observée
q – proportion de non-malades observée (1-p)
$z_{\alpha/2}$ – Percentile de la loi normale selon α/2
n – nombre d'observations dans l'échantillon

Intervalle de confiance exact:
Basé sur la loi binomiale
On veut essayer de déterminer un intervalle regroupant toute valeur de π compatible au niveau α (ou 100α %) avec la proportion observée
π_{sup} est la plus petite des valeurs de π telles que $P(A \leq a \mid \pi) < \alpha/2$
π_{inf} est la plus grande des valeurs de π telles que $P(A \geq a \mid \pi) < \alpha/2$
Calcul itératifs (support informatique incontournable

2 proportions observées

Différence de proportions et Intervalle de confiance asymptotique

$$DP = p_1 - p_0$$

$$DP \pm z_{\alpha/2}\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_0 q_0}{n_0}}$$

p_1 : proportion de malades chez les exposés
p_0 : proportion de malades chez les non exposés
q_1 : proportion de non malades chez les exposés $(1 – p_1)$
q_0 : proportion de non malades chez les non exposés $(1 – p_0)$
n_1 : nombre de sujets exposés
n_0 : nombre de sujets non exposés
$z_{\alpha/2}$: percentile de la loi normale selon α/2

Rapport de proportions et Intervalle de confiance asymptotique

$$RP = \frac{p_1}{p_0}$$

$$RP \times e^{+z_{\alpha/2}\sqrt{\frac{c}{an_1}+\frac{d}{bn_0}}}$$

n_1 : nombre de sujets exposés
n_0 : nombre de sujets non exposés
a : nombre de malades exposés
b : nombre de malades non exposés
c : nombre de non malades exposés
d : nombre de non malades non exposés
$z_{\alpha/2}$: percentile de la loi normale selon α/2

Rapport de cotes et Intervalle de confiance asymptotique

$$RC = \frac{C_1}{C_0}$$

$$RC \times e^{\pm z_{\alpha/2}\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}}$$

a : nombre de malades exposés
b : nombre de malades non exposés
c : nombre de non malades exposés
d : nombre de non malades non exposés
$z_{\alpha/2}$: percentile de la loi normale selon α/2

Tableau de contingence

1 proportion

	Obs	Att
M+	a	nπ
M-	n-a	n(1-π)
	n	

2 proportions observées

Observée (attendue)	E+	E-	
M+	a $(n_1 x p)$	b $(n_0 x p)$	m_1
M-	c $(n_1 x (1-p))$	d $(n_0 x (1-p))$	m_0
	n_1	n_0	N
	$p_1 = a/n_1$ $c_1 = a/c$	$p_0 = b/n_0$ $c_0 = b/d$	$p = m_1/N$ $c = m_1/m_0$

Tests statistique

1 proportion

Valeur-p asymptotique:
- Basé sur la loi normale ou la loi du χ^2
- Facile à calculer
- Approprié seulement si: nombre de cas *et* de non-cas *attendu* ≥5

Valeur-p exacte :
- Basé sur la loi Binomiale
- Complexe à calculer (support informatique presque incontournable)
- Toujours approprié

2 proportions observées

Valeur-p asymptotique (Khi carré de Pearson)
- Basé sur la loi normale ou la loi du $\chi 2$
- Facile à calculer
- Approprié seulement si le nombre de participants attendu ≥ 5 dans chacune des cellules du tableau de contingence

Valeur-p exacte (test exact de Fisher)
- Basé sur la loi hypergéométrique
- Complexe à calculer (support informatique incontournable)
- Test toujours approprié

Carin Ahouada. Médecin Interniste béninois. Etudiant Maitrise en Epidémiologie. Université Laval. 2016-2018

Analyse de proportions		
	1 proportion	**2 proportions observées**
Tests statistique bilatéral	**Test statistique asymptotique bilatéral** $H_0: p = \pi$ $H_1: p \neq \pi$ $\alpha = 0.05$ $Z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$ $\chi^2 = \sum \frac{(O - Att)^2}{Att}$ $Z = \frac{(a - n\pi)}{\sqrt{n\pi(1-\pi)}}$ Valeur-p bilatérale= P(Z≥z) + P(Z≤-z) selon la loi normale ou P(X ≥x) selon la loi du Khi-carré Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p < α, on rejette H_0 **Test statistique exact** $H_0: p = \pi$ $H_1: p \neq \pi$ Si a>E(A) alors valeur-p = P(A≥a) + P(A≤u) avec P(A=u) ≤ P(A=a) Si a<E(A) alors valeur-p = P(A≤a) + P(A≥u) avec P(A=u) ≤ P(A=a) Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p < α, on rejette H_0 n.b. la distribution binomiale n'est pas symétrique sauf si π=0.5	**Test statistique asymptotique bilatéral** H_0: $p_1 = p_0$; DP = 0; RP = 1; RC = 1 H_1: $p_1 \neq p_0$; DP ≠ 0; RP ≠ 1; RC ≠ 1 α = 0,05 Khi carré de Pearson : $\chi_1^2 = \sum \frac{(O - Att)^2}{Att}$ Valeur-p bilatérale = P(X ≥ x) selon la loi du Khi carré Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p < α, on rejette H_0 **Tests statistiques asymptotiques bilatéraux alternatifs** **Loi normale:** $z = \frac{p_1 - p_0}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_0 q_0}{n_0}}}$ **Mantel-Haenszel:** $\chi_1^2 = \frac{(a - E(A))^2}{V(A)}$ **Avec E(A) et V(A) calculées selon la loi hypergéométrique:** $E(A) = \frac{m_1 n_1}{n}$ $V(A) = \frac{m_1 m_0 \; n_1 n_0}{n^2(n-1)}$ **Test statistique exact bilatéral** H_0: $p_1 = p_0$; DP = 0; RP = 1; RC = 1 H_1: $p_1 \neq p_0$; DP ≠ 0; RP ≠ 1; RC ≠ 1 E(A) = a (nombre de malades exposés) Calcul de la valeur-p Si a > E(A) alors valeur-p = P(A ≥ a) + P(A ≤ u) avec P(A = u) ≤ P(A = a) Si a < E(A) alors valeur-p = P(A ≤ a) + P(A ≥ u) avec P(A = u) ≤ P(A = a) Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p < α, on rejette H_0
Test unilatéral à droite	**Test statistique exact:** $H_0: p \leq \pi$ $H_1: p > \pi$ a - nombre de malades observé Valeur-p = P(A≥a) selon la loi Binomiale P(A ≥ a) = $\sum_{x=a}^{n} C_n^x \pi^x (1-\pi)^{n-x}$ Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p < α, on rejette H_0	**Test statistique exact:** H_0: $p_1 \leq p_0$; DP ≤ 0; RP ≤ 1; RC ≤ 1 H_1: $p_1 > p_0$; DP > 0; RP > 1; RC > 1 E(A) = a (nombre de malades exposés) Valeur-p = P(A ≥ a) selon la loi hypergéométrique $P(A \geq a) = \sum_{x=a}^{\min(n_1, m_1)} \frac{C_{n_1}^x \times C_{n_0}^{m-x}}{C_n^m}$ Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p < α, on rejette H_0

Carin Ahouada. Médecin Interniste béninois. Etudiant Maitrise en Epidémiologie. Université Laval. 2016-2018

Analyse de proportions

	1 proportion	2 proportions observées	
Test unilatéral à gauche	**Test statistique exact** $H_0: p \geq \pi$ $H_1: p < \pi$ A=a (nombre de malades observé) Valeur-p = $P(A \leq a)$ selon la loi Binomiale $P(A \leq a) =$ $\sum_{x=0}^{a} C_n^x \pi^x (1-\pi)^{n-x}$ Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p < α, on rejette H_0	**Test statistique exact:** H_0: $p_1 \geq p_0$; DP ≥ 0; RP ≥ 1; RC ≥ 1 H_1: $p_1 < p_0$; DP < 0; RP < 1; RC < 1 E(A) = a (nombre de malades exposés) Valeur-p = $P(A \leq a)$ selon la loi hypergéométrique $P(A \leq a) = \sum_{x=\max(0,n_1-m_0)}^{a} \frac{C_{n_1}^{x} \times C_{n_0}^{m_1-x}}{C_n^{m_1}}$ Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p < α, on rejette H_0	
Conditions	**Estimation d'une proportion** Échantillon est représentative de la population cible Observations indépendantes Calculs exacts utilisé si peu effectifs ou p proche de 0 ou de 1 Absence d'erreur de mesure **Utilisation des tests** Si une des valeurs attendue dans le tableau de contingence <5, on doit utiliser des calculs exacts L'échantillon (population d'étude) est représentatif de la population cible Les observations sont indépendantes Absence d'erreur de mesure	**Estimation DP, RP, RC** Les observations sont indépendantes. Les calculs exacts sont utilisés si peu d'effectifs. L'échantillon est représentatif de la population cible. Absence d'erreur de mesure. Absence de biais de confusion **Utilisation des tests** Les observations sont indépendantes. Si une des valeurs attendues dans le tableau de contingence est < 5, on doit utiliser des calculs exacts. L'échantillon (population d'étude) est représentatif de la population cible. Absence d'erreur de mesure. Absence de biais de confusion	
Taille de l'échantillon	**Pour estimer une proportion p avec une précision i** $n = \frac{z_{\alpha/2}^2 \pi(1-\pi)}{i^2}$ n – nombre d'observations requis $z_{\alpha/2}$ – percentile de la loi normale associé à α/2 π – proportion de malades i – précision désirée (moitié de la largeur de l'intervalle de confiance) **Pour détecter une différence Δ entre une proportion observée et une proportion théorique** $n = \frac{\pi(1-\pi)(z_{\alpha/2}+z_{\beta})^2}{\Delta^2}$ n – nombre d'observations requis $z_{\alpha/2}$ – percentile de la loi normale associé à α/2 (Z_α pour un test unilatérale) z_β – percentile de la loi normale associé à β (1-β est la puissance désirée) π – proportion de malades Δ – Différence entre p et π qu'on souhaite détecter	**selon un RP** $n_1 = \frac{\left(z_{\alpha/2}\sqrt{\frac{(RP+k)[1+k-\pi_0(RP+k)]}{1+k}} + z_\beta\sqrt{kRP(1-RP\pi_0)+(1-\pi_0)}\right)^2}{k\pi_0(RP-1)^2}$ $k = n_0 / n_1$ $n_0 = kn_1$ $n = n_1 + n_0$ n_1 : nombre de sujets exposés requis $z_{\alpha/2}$: percentile de la loi normale associé à α/2 (Z_α pour un test unilatéral) z_β : percentile de la loi normale associé à β (1-β est la puissance désirée) π_0 : proportion de malades chez les non exposés RP : rapport de proportions qu'on souhaite détecter	**selon un RC (étude cas-témoins)** $n_1 = \frac{\left(z_{\alpha/2}\sqrt{\frac{(RC+kA)(1+kA)}{1+k}} + z_\beta\sqrt{kRC+A^2}\right)^2}{k\pi_0(1-\pi_0)[RC-1]^2}$ $A = RC\pi_0 + (1-\pi_0)$ $k = n_0 / n_1$ (nombre de témoins/nombre de cas) $n_0 = kn_1$ $n = n_1 + n_0$ n_1 : nombre de **cas** requis $z_{\alpha/2}$: percentile de la loi normale associé à α/2 (Z_α pour un test unilatéral) z_β : percentile de la loi normale associé à β (1-β est la puissance désirée) π_0 : **proportion d'exposés chez les témoins** RC : rapport de cotes qu'on souhaite détecter N.B. dans une étude cohorte n_1 correspond au nombre de sujets exposés requis et π_0 correspond à la proportion de malades chez les non exposés

Carin Ahouada. Médecin Interniste béninois. Etudiant Maitrise en Epidémiologie. Université Laval. 2016-2018

<table>
<tr><th colspan="5">Analyse de proportions</th></tr>
<tr><th></th><th>1 proportion</th><th colspan="3">2 proportions observées</th></tr>
<tr><td>Puissance</td><td></td><td>selon une différence clinique (RP)
$$z_\beta = \frac{\pi_0(RP-1)}{\sqrt{\pi(1-\pi)}}\sqrt{\frac{n_1 n_0}{n}} - z_{\alpha/2}$$
n_1 : nombre de sujets exposés
n_0 : nombre de sujets non exposés
n : nombre de sujets total
$z_{\alpha/2}$: percentile de la loi normale associé à α/2 (Z_α pour un test unilatéral)
z_β : percentile de la loi normale associé à β (1-β est la puissance désirée)
π_0 : proportion de malades chez les non exposés
π : proportion de malades
RP : rapport de proportions qu'on souhaite détecter</td><td colspan="2">selon une différence clinique (RC)
$$z_\beta = \frac{\pi_0 \frac{(1-\pi_0)}{\pi_0 + 1/(RC-1)}}{\sqrt{\pi(1-\pi)}}\sqrt{\frac{n_1 \times n_0}{n}} - z_{\alpha/2}$$
n_1 : nombre de cas requis
$z_{\alpha/2}$: percentile de la loi normale associé à α/2 (Z_α pour un test unilatéral)
z_β : percentile de la loi normale associé à β (1-β est la puissance désirée)
π_0 : proportion d'exposés chez les témoins
RC : rapport de cotes qu'on souhaite détecter
N.B. dans une étude cohorte π_1 correspond à la proportion de malades chez les exposés et π_0 correspond à la proportion de malades chez les non exposés</td></tr>
<tr><td>SAS</td><td>Test statistique
PROC FREQ;
TABLES y / binomial (level=2 p=π) ;
EXACT binomial;
RUN;

Taille pour detecter une difference
proc power;
onesampleFREQ test=Z method=normal
alpha=0.05
nullproportion= π
sides=2
proportion=p
ntotal= .
power= 1-β;
RUN;</td><td>Estimer la mesure d'association
PROC SORT;
BY m descending e descending;
PROC FREQ order=data;
TABLES e*m / relrisk diffrisk;
RUN;
où e représente la variable désignant l'exposition et m la variable désignant la maladie

Test statistique
PROC FREQ;
TABLES m*e / chisq fisher;
EXACT fisher;
RUN;
où e représente la variable désignant l'exposition et m la variable désignant la maladie</td><td>Taille Selon Rapport de proportions

PROC POWER;
TWOSAMPLEFREQ
test=pchi
refproportion = π₀
groupweights = (w1 w2)
relativerisk = RP
ntotal = .
power = 1-β;
RUN;</td><td>Taille selon Rapport de cotes

PROC POWER;
TWOSAMPLEFREQ
test=pchi
refproportion = π_0
groupweights = (w1 w2)
oddsratio = RC
ntotal = .
power = 1-β;
run;</td></tr>
</table>

Carin Ahouada. Médecin Interniste béninois. Etudiant Maitrise en Epidémiologie. Université Laval. 2016-2018

Test statistique comparant plusieurs proportions	Test statistique comparant plusieurs distributions
Tableau de contingence c x 2 Test global H_0: $p_1 = p_2 = \ldots = p_j$ H_1: au moins deux des proportions diffèrent **Test exact** Toujours approprié Nécessaire si att < 5 dans plus de 20% (1/5) des cellules Loi hypergéométrique multiple (extension du test exact de Fisher) **Test asymptotique:** Khi carré de Pearson: $\chi^2_{c-1} = \sum \frac{(O - Att)^2}{Att}$ Approprié seulement si att > 5 dans plus de 80% (4/5) des cellules	Tableau de contingence c x r Pas de mesure d'association Test global H_0: distribution$_1$ = distribution$_2$ = … = distribution$_j$ H_1: au moins deux des distributions diffèrent Test exact Toujours approprié Nécessaire si att < 5 dans plus de 20% (1/5) des cellules Loi hypergéométrique multiple (extension du test exact de Fisher) Test asymptotique: Khi carré de Pearson: $\chi^2_{(c-1)\times(r-1)} = \sum \frac{(O - Att)^2}{Att}$ Approprié seulement si att > 5 dans plus de 80% (4/5) des cellules

Carin Ahouada. Médecin Interniste béninois. Etudiant Maitrise en Epidémiologie. Université Laval. 2016-2018

<table>
<tr><th colspan="5">Analyse des taux</th></tr>
<tr><th></th><th>1 échantillon</th><th>Un taux observé VS un taux théorique</th><th>2 taux observés</th><th>Plusieurs taux</th></tr>
<tr>
<td>Mesures
De fréquence ou
d'association</td>
<td>Estimation d'un taux/Intervalle de confiance asymptotique
$$IC: \quad t \pm z_{\alpha} \sqrt{\frac{t}{n_t}}$$
t : taux observé
$z_{\alpha/2}$: percentile de la loi normale selon α/2
n : nombre d'unités de personnes-temps dans l'échantillon
Attention! Si la borne inférieure < 0, utilisez calculs exacts!
Estimation d'un taux/Intervalle de confiance exact
On veut essayer de déterminer un intervalle regroupant toutes valeurs de τ compatibles au niveau α (ou 100α %) avec le taux observé
$τ_{sup}$ est la plus petite des valeurs de τ telles que P(A ≤ a | τ) < α/2
$τ_{inf}$ est la plus grande des valeurs de τ telles que P(A ≥ a | τ) < α/2
Basé sur la loi de poisson
Calculs itératifs (support informatique incontournable)</td>
<td></td>
<td>
<table>
<tr><th>Observé(attendu)</th><th>E+</th><th>E-</th><th></th></tr>
<tr><td>M+</td><td>a ($λ_1 = t\,n_1$)</td><td>b ($λ_0 = t\,n_0$)</td><td>m_1</td></tr>
<tr><td>Personne-temps</td><td>n_1</td><td>n_0</td><td>N</td></tr>
<tr><td></td><td>$t_1 = a / n_1$</td><td>$t_0 = b / n_0$</td><td>$t = m_1 / n$</td></tr>
</table>
Différence de taux:
• DT = $t_1 - t_0$
Rapport de taux
• RT = t_1 / t_0
Intervalle de confiance
Exact:
Basé sur la loi binomiale
Toujours approprié
Asymptotique :
Basé sur la loi normale
Seulement approprié si le nombre de cas observés (a et b) ≥ 5
Différence de taux
Intervalle de confiance asymptotique
$$DT = t_1 - t_0$$
$$DT \pm z_{\alpha/2} \sqrt{\frac{t_1}{n_1} + \frac{t_0}{n_0}}$$
t_1 : taux chez les exposés
t_0 : taux chez les non exposés
n_1 : nombre d'unités de personnes-temps d'observation chez les exposés
n_0 : nombre d'unités de personnes-temps d'observation chez les non-exposés
$z_{\alpha/2}$: percentile de la loi normale selon α/2
Rapport de taux
Intervalle de confiance asymptotique
$$RT = t_1 \,/\, t_0$$
$$RT \times e^{\pm z_{\alpha/2} \sqrt{\frac{1}{a} + \frac{1}{b}}}$$
a : nombre de cas exposés
b : nombre de cas non exposés
$z_{\alpha/2}$: percentile de la loi normale selon α/2
</td>
<td></td>
</tr>
</table>

Carin Ahouada. Médecin Interniste béninois. Etudiant Maitrise en Epidémiologie. Université Laval. 2016-2018

Analyse des taux				
	1 échantillon	**Un taux observé VS un taux théorique**	**2 taux observés**	**Plusieurs taux**
Test bilatéral asymptotique		H_0: t = τ H_1: t ≠ τ α = 0,05 $z = \frac{t-\tau}{s_t} = \frac{t-\tau}{\sqrt{\frac{\tau}{n}}}$ ou $\chi^2 = \sum \frac{(O-Att)^2}{Att}$ Valeur-p bilatérale = P(Z ≥ |z|) + P(Z ≤ -|z|) selon la loi normale ou P(X ≥ x) selon la loi du Khi-carré Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p < α, on rejette H_0	H_0: $t_1 = t_0$; DT = 0; RT = 1 H_1: $t_1 \neq t_0$; DT ≠ 0; RT ≠ 1 α = 0,05 $\chi_1^2 = \sum \frac{(O-Att)^2}{Att}$ Valeur-p bilatérale = P(X ≥ x) selon la loi du Khi-carré Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p < α, on rejette H_0	Tableau de contingence c x 1 Test global H_0: $t_1 = t_2 = \ldots = t_j$ H_1: au moins deux des proportions diffèrent **Test exact** Toujours approprié Nécessaire si att < 5 dans plus de 20% (1/5) des cellules Loi multinomiale (extension de la loi binomiale) **Test asymptotique** Khi-carré de Pearson: Approprié seulement si att > 5 dans plus de 80% (4/5) des cellules $\chi_{c-1}^2 = \sum \frac{(O-Att)^2}{Att}$
Test bilatéral exact		H_0: t = τ ou λ = a H_1: t ≠ τ ou λ ≠ a où λ = E(A) Si a > λ alors valeur-p = P(A ≥ a) + P(A ≤ u) avec P(A = u) ≤ P(A = a) Si a < λ alors valeur-p = P(A ≤ a) + P(A ≥ u) avec P(A = u) ≤ P(A = a) Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p < α, on rejette H_0	H_0: $t_1 = t_0$; DT = 0; RT = 1 H_1: $t_1 \neq t_0$; DT ≠ 0; RT ≠ 1 sachant λ = E(A) Si a > λ alors valeur-p = P(A ≥ a) + P(A ≤ u) avec P(A = u) ≤ P(A = a) Si a < λ alors valeur-p = P(A ≤ a) + P(A ≥ u) avec P(A = u) ≤ P(A = a) Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p > α, on rejette H_0	
Test exact unilatéral à droit		H_0: t ≤ τ H_1: t > τ Valeur-p = P(A ≥ a) selon la loi de Poisson $P(A \geq a \mid \lambda) = \sum_{x=a}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!}$ Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p < α, on rejette H_0	H_0: $t_1 \leq t_0$; DT ≤ 0; RT ≤ 1 H_1: $t_1 > t_0$; DT > 0; RT > 1 Valeur du test statistique = a (nombre de malades exposés) Valeur-p = P(A ≥ a) selon la loi binomiale $P(A \geq a) = \sum_{x=a}^{m} C_m^x \pi^x (1-\pi)^{m-x}$ où $\pi = \frac{n_1}{n_1+n_0}$ Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p > α, on rejette H_0	
Test exact unilatéral à gauche		H_0: t ≥ τ H_1: t < τ Valeur-p = P(A ≤ a) selon la loi de Poisson $P(A \leq a \mid \lambda) = \sum_{x=0}^{a} \frac{e^{-\lambda}\lambda^x}{x!}$ Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p < α, on rejette H_0	H_0: $t_1 \geq t_0$; DT ≥ 0; RT ≥ 1 H_1: $t_1 < t_0$; DT < 0; RT < 1 Valeur du test statistique = a (nombre de malades exposés) Valeur-p = P(A ≤ a) selon la loi binomiale $P(A \leq a) = \sum_{x=a}^{m} C_m^x \pi^x (1-\pi)^{m-x}$ où $\pi = \frac{n_1}{n_1+n_0}$ Si la valeur-p ≥ α, on accepte H_0 Si la valeur-p > α, on rejette H_0	

Carin Ahouada. Médecin Interniste béninois. Etudiant Maitrise en Epidémiologie. Université Laval. 2016-2018

Conditions	Les unités personnes-temps sont indépendantes. Les calculs exacts sont utilisés si peu d'effectifs. La population est stable dans le temps. L'échantillon est représentatif de la population cible. Absence d'erreur de mesure. Absence de biais de confusion.		Les unités de personnes-temps sont indépendantes. Si une des valeurs attendue dans le tableau de contingence est < 5, on doit utiliser des calculs exacts. L'échantillon (population d'étude) est représentatif de la population cible. La population (taux) est stable dans le temps. Absence d'erreur de mesure. Absence de biais de confusion	
Taille population		selon une précision i $n = \frac{z_{\alpha/2}^{2}\tau}{i^{2}}$ n : nombre de personnes-temps d'observation requis $z_{\alpha/2}$: percentile de la loi normale associé à α/2 τ : taux i : précision désirée (moitié de la largeur de l'intervalle de confiance)	**selon un RT** $n_1 = \frac{\left(z_{\alpha/2}\sqrt{RT+k} + z_{\beta}\sqrt{kRT+1}\right)^2}{k\tau_0(RT-1)^2}$ $k = n_0 / n_1$ $n_0 = kn_1$ $n = n_1 + n_0$ n_1 : nombre d'unités de personnes-temps exposées requises $z_{\alpha/2}$: percentile de la loi normale associé à α/2 (Z_α pour un test unilatérale) z_β : percentile de la loi normale associé à β (1-β est la puissance désirée) RT : rapport de taux qu'on souhaite détecter **selon un RT (ou une DT)** $z_\beta = \frac{\lvert\Delta\rvert}{\sqrt{\tau}}\sqrt{\frac{n_1 n_0}{n}} - z_{\alpha/2}$ z_β : percentile de la loi normale associé à β (1-β est la puissance désirée) Δ : différence de taux $\Delta = \tau_0(RT-1)$ n_0 : nombre d'unités de personnes-temps non-exposés n_1 : nombre d'unités de personnes-temps exposés n : nombre d'unités de personnes-temps total $z_{\alpha/2}$: percentile de la loi normale associé à α/2 (z_α pour un test unilatérale)	
Sas			data probtaux; *valeur-p exacte unilatérale à droite pour un taux observé; P = 1-POISSON (λ, a); *valeur-p asymptotique bilatérale pour un taux observé; P = 1-PROBCHI(χ^2, ddl); *valeur-p exacte unilatérale à droite pour deux taux observés; P = 1-PROBBNML(n, π, a); *valeur-p asymptotique bilatérale pour deux taux observés; P = 1-PROBCHI(χ^2, ddl); proc print;run;	

Carin Ahouada. Médecin Interniste béninois. Etudiant Maitrise en Epidémiologie. Université Laval. 2016-2018

<table>
<tr><th colspan="3">Analyse de données appariées</th></tr>
<tr><td>Tableau de contingence</td><td colspan="2">Organisation des données par paire
<table>
<tr><td></td><td></td><td colspan="2">Témoins</td></tr>
<tr><td></td><td></td><td>E+</td><td>E-</td></tr>
<tr><td rowspan="2">Cas</td><td>E+</td><td>f_{11}</td><td>f_{10}</td></tr>
<tr><td>E-</td><td>f_{01}</td><td>f_{00}</td></tr>
</table>
f_{11} : Nombre de paires où le cas et le témoin sont exposés
f_{10} : Nombre de paires où seulement le cas est exposé
f_{01} : Nombre de paires où seulement le témoin est exposé
f_{00} : Nombre de paires où le cas et le témoin sont non-exposés
$RC = \frac{f_{10}}{f_{01}}$</td></tr>
<tr><td>Intervalle de confiance asymptotique pour le rapport de cotes</td><td colspan="2">$RC \times e^{\pm z_{\alpha/2}\sqrt{V(\log RC)}}$
$S_{\log RC} = \sqrt{V(\log RC)} \approx \sqrt{\frac{f_{10}+f_{01}}{f_{10}f_{01}}}$</td></tr>
<tr><td>Intervalle de confiance exact pour le rapport de cotes</td><td colspan="2">La limite inférieure ψ_i est telle que $P(A \geq f_{10} | \psi_i) = \alpha/2$
La limite supérieure ψ_s est telle que $P(A \leq f_{10} | \psi_s) = \alpha/2$
A : nombre de paires où seulement le cas est exposé
$A \mapsto$ Binomiale[n, $\pi(\psi)$] où $n = f_{10}+f_{01}$ et $\pi(\psi) = \psi/(1+\psi)$</td></tr>
<tr><td>Interprétation</td><td colspan="2">L’intervalle de confiance correspond à l’intervalle à l’intérieur duquel on est confiant à (1-α)X100% que le vrai RC de la population cible se trouve
L’intervalle de confiance indique la précision de notre estimation
On peut utiliser l’intervalle de confiance pour tester H_0: $\psi = 1$</td></tr>
<tr><td>Conditions</td><td colspan="2">Les calculs exacts devraient être utilisés si peu d’effectifs ($f_{10} < 5$ ou $f_{01} < 5$).
L’échantillon est représentatif de la population cible.
Absence d’erreur de mesure. Absence de biais de confusion.
Témoins sélectionnés dans la population qui a donné naissance aux cas</td></tr>
<tr><td>Test statistique</td><td>Calculs asymptotiques
• Test de McNemar
• Test basé sur la loi du χ^2
• Facile à calculer
• Seulement appropriés si $E(f_{10}) \geq 5$</td><td>Calculs exacts
• Basés sur la loi binomiale
• Complexes (support informatique incontournable)
• Toujours appropriés</td></tr>
<tr><td>Test statistique asymptotique bilatéral
(Test de McNemar)</td><td colspan="2">1. H_0: $\psi = 1$
H_1: $\psi \neq 1$
2. α = 0,05
3.
$\chi_1^2 = \frac{(f_{10} - f_{01})^2}{f_{10} + f_{01}}$</td></tr>
</table>

Carin Ahouada. Médecin Interniste béninois. Etudiant Maitrise en Epidémiologie. Université Laval. 2016-2018

	4. Valeur-p bilatérale = P(X ≥ x) selon la loi du Khi-carré 5. Si la valeur-p < α, on rejette H_0	
Test statistique exact unilatéral (à droite)	1. H_0: ψ ≤ 1 H_1: ψ > 1 2. Valeur du test statistique = f_{10} 3. Valeur-p = P(A ≥ a) selon la loi binomiale 4. Si la valeur-p < α, on rejette H_0	$P(A \geq f_{10}) = \sum_{x=f_{10}}^{n} C_n^x \pi^x (1-\pi)^{n-x}$ où $\pi = \frac{\psi}{1+\psi}$ et $n = (f_{10} + f_{01})$
Interprétation	Si on accepte H_0 (valeur-p ≥ α), on conclut que le RC = 1 ou que le RC (ou l'association entre E et M) n'est pas statistiquement significative, qu'elle peut être expliquée par les fluctuations du hasard.	Si on rejette H_0 (valeur-p < α), on conclut que le RC ≠ 1 ou que le RC (ou l'association entre E et M) est statistiquement significative, qu'elle ne peut pas être expliquée par les fluctuations du hasard.
Calcul de taille d'échantillon pour détecter un RC de ψ	On aura besoin d'au moins n paires discordantes pour qu'un RC de ψ soit statistiquement significatif $n = \frac{\left[z_{\alpha/2}(\psi+1) + 2z_\beta\sqrt{\psi}\right]^2}{(\psi-1)^2}$	n : nombre de paires discordantes requis $z_{\alpha/2}$: percentile de la loi normale associé à α/2 ψ : estimation du RC (littérature ou étude pilote) z_β : percentile de la loi normale pour β
Dans le logiciel SAS	**DATA** exemple; INPUT f1 f2 n; DATALINES; 1 1 6 1 0 5 0 1 8 0 0 10 ; **RUN**; f_1 : exposition chez les cas f_2 : exposition chez les	**PROC FREQ** DATA = exemple; TABLES *f1***f2* / AGREE; **RUN**;

Carin Ahouada. Médecin Interniste béninois. Etudiant Maitrise en Epidémiologie. Université Laval. 2016-2018

Analyse de tendance		
Analyse de tendance	**Contexte :** • Plus de 2 proportions, cotes ou taux • Exposition mesurée sur une échelle ordinale • Plausibilité biologique d'une association linéaire	**Démarche :** • Tableau de contingence • Calcul des mesures de fréquence (p, t, ou c) et des mesures d'association (RP, RT, ou RC) • Représentation graphique pour évaluer la tendance linéaire • Test statistique
Test statistique	Asymptotique : • Test du χ^2 de Mantel-Haenszel • Test d'Armitage-Cochrane (asymptotiquement équivalent) Approprié seulement si ≥5 observations attendues dans au moins 80% des cellules	
Test statistique Test du χ^2 de Mantel-Haenszel	$\chi_1^2 = \frac{\left[\sum_j a_j X_j - \frac{a}{n}\sum_j n_j X_j\right]^2}{V\left(\sum_j a_j X_j\right)}$ $V\left(\sum_j a_j X_j\right) = \frac{ab}{n^2(n-1)}\left[n\sum_j n_j X_j^2 - \left(\sum_j n_j X_j\right)^2\right]$	X : valeur au centre de la catégorie de la variable indépendante a : nombre de cas b: nombre de non-cas n : nombre total
Interprétation	Si on accepte H_0 (valeur-p ≥ α), on conclut que les proportions ne suivent pas une tendance linéaire, qu'il n'y a pas d'association **linéaire** significative entre les deux variables.	Si on rejette H_0 (valeur-p < α), on conclut que les proportions varie de manière linéaire avec les catégories d'exposition, qu'il y a association significative entre les deux variables.
Conditions	• Les observations sont indépendantes • Si > 20% des valeurs attendues dans le tableau de contingence sont < 5, on doit utiliser des calculs exacts • Une association linéaire est scientifiquement plausible	• L'échantillon (population d'étude) est représentatif de la population cible • Absence d'erreur de mesure • Absence de biais de confusion

Carin Ahouada. Médecin Interniste béninois. Etudiant Maitrise en Epidémiologie. Université Laval. 2016-2018

Types de régression

Modèle	**Échelle**	**Modèle**	**Mesure d'association**	**Méthode d'estimation des coefficients**
Linéaire	Additive	$\mu = \alpha + \beta X$	Différence de Moyennes (DM) Rapport de cotes	Moindres carrés
Logistique	Multiplicative	$\log \frac{\pi}{1-\pi} = \alpha + \beta X$	Rapport de Cotes (RC)	Maximum de Vraisemblance
Binomiale	Additive Multiplicative	$\pi = \alpha + \beta X$ $\log \pi = \alpha + \beta X$	Différence de Proportions (DP) Rapport de proportions (RP)	Maximum de Vraisemblance
Poisson	Additive Multiplicative	$\lambda = \alpha + \beta X$ $\log \lambda = \alpha + \beta X$	Différence de Taux (DT) Rapport de Taux (RT)	Maximum de Vraisemblance
Cox	Multiplicative	$\log \frac{\lambda(t)}{\lambda_0(t)} = \beta X$	Rapport de Taux (RT)	Maximum de vraisemblance

Carin Ahouada. Médecin Interniste béninois. Etudiant Maitrise en Epidémiologie. Université Laval. 2016-2018